A CONSTRAINED OPTIMIZATION APPROACH FOR COMPLEX SPARSE PERTURBED MODELS[*]

ANISIA FLORESCU[†], EMILIE CHOUZENOUX[‡], JEAN-CHRISTOPHE PESQUET[§], AND SILVIU CIOCHINA[¶]

Abstract. In this paper, we consider the problem of estimating a complex-valued signal having a sparse representation in an uncountable family of vectors. The available observations are corrupted with an additive noise and the elements of the dictionary are parameterized by a scalar real variable. By a linearization technique, the original model is recast as a constrained sparse perturbed model. An optimization approach is then proposed to estimate the parameters involved in this model. The cost function includes an arbitrary Lipschitz differentiable data fidelity term accounting for the noise statistics, and an ℓ_0 penalty. A forward-backward algorithm is employed to solve the resulting non-convex and non-smooth minimization problem. This algorithm can be viewed as a generalization of an iterative hard thresholding method and its local convergence can be established. Simulation results illustrate the good practical performance of the proposed approach when applied to spectrum estimation.

Key words. complex-valued data, hard thresholding, nonconvex optimization, nonsmooth analysis, linear modelling, perturbations, proximal methods, sparse approximation, signal processing, spectral estimation.

AMS subject classifications. 65K10, 90C26, 94A12, 62M15, 65F22

1. Problem statement. We consider a family of vectors $\mathcal{E} = \{e_\nu \mid \nu \in \mathbb{R}\}$ of $\mathbb{C}^Q$ which are parameterized by a scalar variable $\nu \in \mathbb{R}$. We assume that a signal $\overline{x} \in \mathbb{C}^Q$ admits a sparse representation on a finite subset $\{e_{\overline{\nu}_n}, 1 \leq n \leq N\}$ of distinct elements of $\mathcal{E}$:

$$\overline{x} = \sum_{n=1}^{N} \overline{c}_n e_{\overline{\nu}_n} \tag{1.1}$$

where a large number of components of $\overline{c} = (\overline{c}_n)_{1\leq n\leq N} \in \mathbb{C}^N$ are assumed to be equal to zero. (Overlined variables are used here to distinguish “true” vectors from generic variables). A classical problem in sparse estimation [4] is then to recover $\overline{x}$ from a vector of observations

$$y = \overline{x} + w \tag{1.2}$$

where $w \in \mathbb{C}^Q$ is a realization of a random noise vector.

In this work, we will be interested in the case when the parameters $(\overline{\nu}_n)_{1\leq n\leq N}$ are known in an imprecise manner, i.e. they are such that, for every $n \in \{1, \ldots, N\}$,

$$\overline{\nu}_n = \theta_n + \overline{\delta}_n \tag{1.3}$$

[*]Part of this work was supported by PhD Fellowship “Investitii in cercetare-inovare-dezvoltare pentru viitor (DocInvest)”, EC project POSDRU/107/1.5/S/76813.

[†]Politehnica University of Bucharest, Telecommunications Dept., Romania (anisia.florescu@ugal.ro)

[‡]Université Paris-Est, LIGM, UMR CNRS 8049, 77454 Marne-la-Vallée, France (emilie.chouzenoux@univ-mlv.fr)

[§]Université Paris-Est, LIGM, UMR CNRS 8049, 77454 Marne-la-Vallée, France (pesquet@univ-mlv.fr)

[¶]Politehnica University of Bucharest, Telecommunications Dept., Romania (silviu@comm.pub.ro)

where $\theta_n \in \mathbb{R}$ is some given value and $\overline{\delta}_n \in \mathbb{R}$ is an unknown error on the parameter to be estimated. If we assume that the perturbations $(\overline{\delta}_n)_{1\le n\le N}$ are small and $\nu \mapsto e_\nu$ is a differentiable function, we can perform the following first-order Taylor expansion:

$$(\forall n \in \{1,\dots,N\}) \qquad e_{\overline{\nu}_n} \simeq e_{\theta_n} + \overline{\delta}_n e'_{\theta_n} \tag{1.4}$$

where e'_{θ_n} is the gradient of $\nu \mapsto e_\nu$ at θ_n. With this approximation, Model (1.1) takes the following bilinear form

$$\overline{x} = \sum_{n=1}^{N} \big(\overline{c}_n e_{\theta_n} + \overline{c}_n \overline{\delta}_n e'_{\theta_n}\big). \tag{1.5}$$

A similar sparse approach for decomposing a signal in terms of translated versions of some features in a finite dictionary is addressed in [7] where the proposed convex ℓ_1 formulation is only applicable to real-valued signals. Likewise, our work can be seen as having some similarities with the perturbed compressive sampling approach in [9] where a robust total least squares (TLS) approach based on an ℓ_1 regularization is developed. The difference is that, in our work, we adopt a different formulation where

- an ℓ_0 cost is employed for the minimization process, instead of its ℓ_1 convex relaxation;
- the errors $(\overline{\delta}_n)_{1\le n\le N}$ are constrained to satisfy the following inequalities
$$(\forall n \in \{1,\dots,N\}) \qquad |\overline{\delta}_n| \le \Delta_n, \tag{1.6}$$
the upper bounds $(\Delta_n)_{1\le n\le N} \in [0,+\infty)^N$ being set by the user. These constraints provide more flexibility than the TLS approach for controling the perturbations $(\overline{\delta}_n)_{1\le n\le N}$;
- a general Lipschitz differentiable data fidelity term can be considered.

Notation. In the following, χ_S denotes the characteristic function of a set S, which is equal to 0 on S and 1 elsewhere, and ι_S denotes the indicator function of a set S, wich is equal to 0 on S and $+\infty$ elsewhere. The transconjugate operation for complex-valued matrices is denoted by $(\cdot)^H$.

2. Proposed optimization approach.

2.1. Variational formulation. We propose to estimate the parameters of the perturbed sparse model by solving the following optimization problem:

$$\underset{\substack{c=(c_n)_{1\le n\le N}\in\mathbb{C}^N\\ \delta=(\delta_n)_{1\le n\le N}\in B}}{\text{minimize}} \;\Phi\Big(\sum_{n=1}^{N}\big(c_n e_{\theta_n} + c_n\delta_n e'_{\theta_n}\big) - y\Big) + \lambda \ell_0(c) \tag{2.1}$$

where B is the N-dimensional box $[-\Delta_1,\Delta_1]\times\cdots\times[-\Delta_N,\Delta_N]$, $\Phi\colon \mathbb{C}^Q \to \mathbb{R}$ is the data-fidelity term which is often chosen equal to the neg-log-likelihood of the noise corrupting the observations, and $\lambda \in (0,+\infty)$ is a regularization constant.

Let us now define the matrices $E = [e_{\theta_1}\dots e_{\theta_N}] \in \mathbb{C}^{Q\times N}$, $E' = [e'_{\theta_1}\dots e'_{\theta_N}] \in \mathbb{C}^{Q\times N}$, and let us introduce the variable $d = (c_n\delta_n)_{1\le n\le N} \in \mathbb{C}^N$. In addition, let the function Ψ be defined as

$$(\forall c = (c_n)_{1\le n\le N} \in \mathbb{C}^N)(\forall d = (d_n)_{1\le n\le N} \in \mathbb{C}^N) \quad \Psi(c,d) = \sum_{n=1}^{N} \psi_n(c_n,d_n) \tag{2.2}$$

where, for every $n \in \{1,\dots,N\}$,

$$(\forall (c_n,d_n)\in\mathbb{C}^2)\qquad \psi_n(c_n,d_n)=\lambda\,\chi_{\{0\}}(c_n)+\iota_{S_n}(c_n,d_n), \tag{2.3}$$

and S_n is the closed cone given by

$$S_n=\{(c_n,d_n)\in\mathbb{C}^2 \mid \exists \delta_n\in[-\Delta_n,\Delta_n], d_n=\delta_n c_n\}. \tag{2.4}$$

Then, Problem (2.1) is equivalent to minimizing function

$$(c,d)\mapsto \Phi\Big([E\ \ E']\begin{bmatrix}c\\ d\end{bmatrix}-y\Big)+\Psi(c,d). \tag{2.5}$$

We have now to see how this minimization can be performed numerically.

2.2. Proposed algorithm. If we assume that Φ is a differentiable function, the previous split form of the objective function suggests the use of a forward-backward algorithm [6]:

$$\begin{array}{l}
(c^{(0)},d^{(0)})\in(\mathbb{C}^N)^2\\
(\underline{\gamma},\overline{\gamma})\in(0,+\infty)^2\\
\text{For }\ k=0,1,\dots\\
\left\lfloor
\begin{array}{l}
\gamma^{(k)}\in(\underline{\gamma},\overline{\gamma})\\
D^{(k)}=\nabla\Phi\Big([E\ \ E']\begin{bmatrix}c^{(k)}\\ d^{(k)}\end{bmatrix}-y\Big)\\
(\widetilde{c}_n^{(k)})_{1\le n\le N}=c^{(k)}-\gamma^{(k)}E^H D^{(k)}\\
(\widetilde{d}_n^{(k)})_{1\le n\le N}=d^{(k)}-\gamma^{(k)}(E')^H D^{(k)}\\
(c_n^{(k+1)},d_n^{(k+1)})_{1\le n\le N}=\big(\operatorname{prox}_{\gamma^{(k)}\psi_n}(\widetilde{c}_n^{(k)},\widetilde{d}_n^{(k)})\big)_{1\le n\le N}.
\end{array}
\right.
\end{array} \tag{2.6}$$

We recall that the proximity operator of a proper, lower bounded, lower semi-continuous function $\varphi\colon \mathcal{H}\to(-\infty,+\infty]$ where $\mathcal{H}$ is a finite dimensional Hilbert space is defined as

$$(\forall u\in\mathcal{H})\qquad \operatorname{prox}_\varphi(u)\in\underset{v\in\mathcal{H}}{\arg\min}\ \frac{1}{2}\|u-v\|^2+\varphi(v). \tag{2.7}$$

Note that, although the uniqueness of $\operatorname{prox}_\varphi(u)$ is guaranteed when φ is convex, this property is not necessarily satisfied in the nonconvex case.
We then have the following result the proof of which is skipped.

Proposition 2.1. *Let $\gamma\in(0,+\infty)$. For every $n\in\{1,\dots,N\}$ and $(c_n,d_n)\in\mathbb{C}^2$, the proximity operator of $\gamma\psi_n$ is*

$$\operatorname{prox}_{\gamma\psi_n}(c_n,d_n)=\begin{cases}(0,0) & \textit{if } |c_n|^2+|d_n|^2<\frac{|\widehat{\delta}_n c_n-d_n|^2}{1+\widehat{\delta}_n^2}+2\gamma\lambda\\[4pt] \dfrac{c_n+\widehat{\delta}_n d_n}{1+\widehat{\delta}_n^2}(1,\widehat{\delta}_n) & \textit{otherwise,}\end{cases} \tag{2.8}$$

where

$$\widehat{\delta}_n=\begin{cases}\min\big\{\frac{\eta_n+|d_n|^2-|c_n|^2}{2|\mathrm{Re}(c_n d_n^*)|},\Delta_n\big\}\operatorname{sign}\big(\mathrm{Re}(c_n d_n^*)\big) & \textit{if } \mathrm{Re}(c_n d_n^*)\neq 0\\ 0 & \textit{if } \mathrm{Re}(c_n d_n^*)=0\\ & \quad\textit{and } |c_n|\ge|d_n|\\ \Delta_n & \textit{otherwise,}\end{cases} \tag{2.9}$$

and $\eta_n = \sqrt{\big(|d_n|^2 - |c_n|^2\big)^2 + 4\big(\mathrm{Re}(c_n d_n^*)\big)^2}$.

It can be noticed that, if $\Delta_n = 0$ for every $n \in \{1,\dots,N\}$, then (2.8) simplifies into

$$\mathrm{prox}_{\gamma\psi_n}(c_n,d_n) = \begin{cases} (0,0) & \text{if } |c_n| < \sqrt{2\gamma\lambda} \\ (c_n,0) & \text{otherwise.} \end{cases} \tag{2.10}$$

This shows that in the absence of perturbations, Algorithm (2.6) reduces to an iterative hard thresholding algorithm [2], when Φ is the squared Euclidean norm. Note however that a main difference between our algorithm and iterative hard thresholding is that, in the unthresholded case, the former one does not reduce to the Landweber algorithm. This makes the analysis of the proposed forward-backward algorithm more difficult.

2.3. Convergence result. As Ψ is a nonconvex function, the convergence of the proposed algorithm requires special care. The following local convergence property can however be proved by using recent results in nonsmooth/nonconvex optimization [1] concerning functions of real variables:

PROPOSITION 2.2. *Assume that Φ is a semi-algebraic function having an L-Lispchitzian gradient with $L \in (0,+\infty)$ and the bounds $\underline{\gamma}$ and $\overline{\gamma}$ on the step-size are chosen such that*

$$0 < \underline{\gamma} \le \overline{\gamma} < L^{-1}\|EE^H + E'(E')^H\|^{-1}. \tag{2.11}$$

Then, any bounded sequence $(c^{(k)}, d^{(k)})_{k\in\mathbb{N}}$ generated by Algorithm (2.6) *converges to a critical point of Function* (2.5).

We recall that a function is semi-algebraic if its graph can be expressed as a finite union of subsets defined by a finite number of polynomial inequalities (in the real and imaginary parts of its complex variables). This includes the standard least squares criterion as a special case.

3. Application to spectrum analysis. Much attention has been paid recently to sparse models in spectrum estimation [3, 8] but, in many methods, an important problem remains the choice of an appropriate search frequency grid.

In this part, we consider an analog complex-valued signal which is expressed as

$$(\forall t \in \mathbb{R}) \qquad s(t) = \sum_{n=1}^{N} \overline{a}_n \exp(\imath \overline{\nu}_n t) \tag{3.1}$$

where $(\overline{\nu}_n)_{1\le n\le N} \in [0,2\pi)^N$ are distinct angular frequencies and $(\overline{a}_n)_{1\le n\le N}$ are unknown complex amplitude values which are assumed to be sparse. This signal is filtered by a known stable continuous-time filter with frequency response G, so yielding

$$(\forall t \in \mathbb{R}) \qquad z(t) = \sum_{n=1}^{N} G(\overline{\nu}_n)\overline{a}_n \exp(\imath \overline{\nu}_n t) + v(t) \tag{3.2}$$

where v is a realization of an additive random noise further corrupting the data. Discrete-time observations are obtained by sampling this signal at distinct times

$(\tau_q)_{1\le q\le Q}$ in $[0,Q]$ in a possibly irregular manner. This leads to Model (1.1)-(1.2) where $y = \big(z(\tau_q)\big)_{1\le q\le Q}$, $w = \big(v(\tau_q)\big)_{1\le q\le Q}$, and $(\overline{c}_n) = \big(G(\overline{\nu}_n)\overline{a}_n\big)_{1\le n\le N}$. In this example, the dictionary consists of cisoids:

$$(\forall \nu \in \mathbb{R}) \qquad e_\nu = \big(\exp(\imath\nu\tau_q)\big)_{1\le q\le Q} \tag{3.3}$$

and we have

$$(\forall \nu \in \mathbb{R}) \qquad e'_\nu = \big(\imath\tau_q\exp(\imath\nu\tau_q)\big)_{1\le q\le Q}. \tag{3.4}$$

A standard choice in this context is to uniformly sample the frequency domain:

$$(\forall n \in \{1,\dots,N\}) \qquad \theta_n = 2\pi\frac{n-1}{N}. \tag{3.5}$$

In order to account for frequencies which are not multiple of $2\pi/N$, the proposed perturbed sparse estimation technique can be applied by choosing, for every $n \in \{1,\dots,N\}$, $\Delta_n = \pi/N$.

As an illustration of the good performance of the proposed approach, we consider $Q = 50$ observations of a complex-valued signal corresponding to the sum of 4 noisy cisoids which have been irregularly sampled in a random manner over $[0,Q]$. The discrete-time observations are corrupted with a white circular Gaussian noise with zero-mean and variance σ^2. Various values of the signal-to-noise ratio have been tested. The employed dictionary consists of $N = 500$ cisoids. The frequencies of the sparse components are not on the search grid. The global normalized root mean square estimation errors for the signal estimated with Algorithm (2.6), when Φ is the squared Euclidean norm, is provided in Table 3.1. These values are the results of a Monte Carlo study performed on 100 realizations. For comparison, we also give the error values corresponding to the use of an ℓ_1 norm and an ℓ_0 cost. The ℓ_1-based solution corresponds to the following convex optimization problem:

$$\underset{c\in S'}{\text{minimize}}\ \ \ell_1(c) \tag{3.6}$$

where $S' = \{c \in \mathbb{C}^N \mid \|Ec - y\|^2 \le Q\sigma^2\}$. A primal-dual proximal algorithm [5] was implemented to efficiently solve this problem. The so-obtained result was chosen as an initial value for the iterative hard thresholding approach associated with the basic ℓ_0 penalized problem. The ℓ_1-based solution was also used as a starting point of our algorithm. Note that a procedure was devised in order to automatically set the regularization parameter λ from the observed data.

The results shown in Figures 3.1 and 3.2 allow us to evaluate the good quality of the estimates typically obtained when identifying 4 or 6 cisoids.

SNR (dB)	ℓ_1	ℓ_0	Proposed method
14.82	0.452	0.224	**0.043**
19.82	0.445	0.206	**0.024**
24.82	0.443	0.198	**0.013**
29.72	0.442	0.201	**0.007**

Table 3.1

Average value of the normalized root mean square reconstruction error computed over 100 realizations.

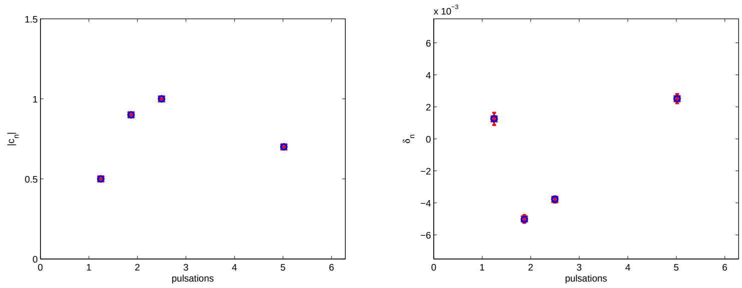

Fig. 3.1. *Estimation results in the case of* 4 *cisoids (SNR* = 19.82 *dB): values of* $(|c_n|)_{1\leq n\leq N}$ *(left); values of frequency perturbations* $(\delta_n)_{1\leq n\leq N}$ *as a function of* $(\theta_n)_{1\leq n\leq N}$ *(right). The exact values are depicted with blue circles and the confidence intervals on the estimates in red (the mean is indicated by a cross).*

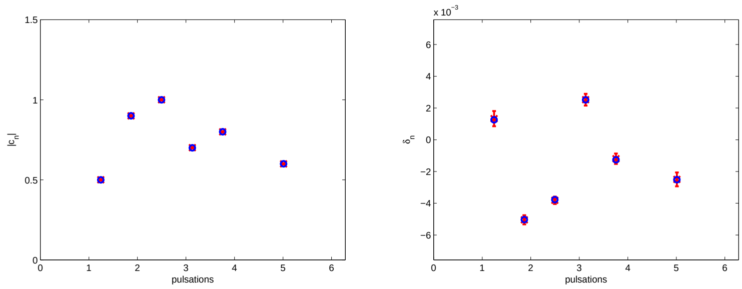

Fig. 3.2. *Estimation results in the case of* 6 *cisoids (SNR* = 23.27 *dB): values of* $(|c_n|)_{1\leq n\leq N}$ *(left); values of frequency perturbations* $(\delta_n)_{1\leq n\leq N}$ *as a function of* $(\theta_n)_{1\leq n\leq N}$ *(right). The exact values are depicted with blue circles and the confidence intervals on the estimates in red (the mean is indicated by a cross).*

REFERENCES

[1] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized Gauss-Seidel methods. *Math. Prog.*, 137(1-2):1–39, Feb. 2013.

[2] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, 2008.

[3] S. Bourguignon, H. Carfantan, and J. Idier. A sparsity-based method for the estimation of spectral lines from irregularly sampled data. *IEEE J. Sel. Topics Signal Process.*, 1(4):575–585, Dec. 2007.

[4] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, Mar. 2008.

[5] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1):120–145, 2011.

[6] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2011.

[7] C. Ekanadham, D. Tranchina, and E. P. Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE Trans. Signal Process.*, 59(10):4735–4744, Oct. 2011.

[8] P. Stoica and P. Babu. Sparse estimation of spectral lines: Grid selection problems and their solutions. *IEEE Trans. Signal Process.*, 60(2):962–967, Feb. 2012.

[9] H. Zhu, G. Leus, and G. B. Giannakis. Sparsity-cognizant total least-squares for perturbed compressive sampling. *IEEE Trans. Signal Process.*, 59(5):2002–2016, May 2011.